PROBABILITY INEQUALITIES AND ERRORS IN CLASSIFICATION*

by

Somesh Das Gupta**

Technical Report No. 190

November 1972

University of Minnesota
Minneapolis, Minnesota

*American Mathematical Society 1970 Subject Classification: Primary 62H30; Secondary 60E05.

Key words and phrases: Probability inequalities; multivariate normal distribution; classification; two populations; probability of correct classification; monotonicity; estimates of probability or correct classification.

**This research was supported by U.S. Army Grant DA-ARO-D-31-124-70-G-102.

1. Introduction.

Let X and Y be two $p \times 1$ random vectors, distributed according to the 2p-variate normal distribution with mean vectors μ and $a\mu$, respectively, and covariance matrix

$$(1.1) \qquad \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \otimes I_p .$$

Let S be a random $p \times p$ matrix distributed as the Wishart distribution $W_p(I_p, f)$, independently of X and Y. Let

$$(1.2) \qquad G(\mu; a, \rho, c) = P[X'Y < c],$$

$$(1.3) \qquad H(\mu; a, \rho, c) = P[X'S^{-1}Y < c].$$

It may be noted that

$$(1.4) \qquad G(\mu; a, \rho, c) = 1 - G(\mu; -a, -\rho, -c)$$

and the same relation holds for H.

In Section 2 the following probability inequalities are proved.

Theorem 1.

The function G involves μ through $\mu'\mu$ and it is a monotonic increasing function of $\mu'\mu$, if any of the following conditions hold:

(i) $|\rho| < 1$; $a = -1$, or $-1 < a \leq 0$, $c \leq 0$.

(ii) $|\rho| < 1$; $a < -1$, $c \leq 0$.

(iii) $\rho = 1$; $a = -1$, or $a \leq 0$, $a \neq -1$, $c \leq 0$. $\rho = -1$, $a = -1$.

G *is a strictly increasing function of* $\mu'\mu$ *unless* $a = c = \rho = 0$, *in* (i), *and* $c + (1-a)^2\mu'\mu/4 < 0$, $\rho = 1$ or $\rho = -1$, $a = -1$, $c \geq 0$ *in* (iii).

Thoerem 2.

The function H is a monotonic increasing function of $\mu'\mu$, if any of the conditions (i)-(iii) of Theorem 1 holds; it is a strictly increasing function of $\mu'\mu$ unless $a = c = \rho = 0$ *in* (i), *or* $\rho = -1$, $a = -1$, $c \geq 0$.

These results are used in Section 3 to prove a monotonicity property of the probabilities of correct classification of a class of rules for classification into two multivariate normal populations. Let X, X_1, X_2 be three mutually independent random $p \times 1$ vectors distributed as $N_p(\mu, \Sigma)$, $N_p(\mu_1, \Sigma/a_1)$, and $N_p(\mu_2, \Sigma/a_2)$, respectively. Let S be a random matrix distributed as $W_p(\Sigma, f)$, independently of X, X_1, and X_2. The problem is to decide whether $\mu = \mu_1$ or $\mu = \mu_2$; a_1 and a_2 are known constants.

When Σ is known (and taken to be I_p), we consider a class of decision rules given by $\varphi_k (k > 0, c \leq 0)$ which decides $\mu = \mu_1$, iff

$$k(1 + 1/a_1)^{-1}\|X - X_1\|^2 < (1 + 1/a_2)^{-1}\|X - X_2\|^2 + c. \tag{1.5}$$

When Σ is unknown, we consider a class of decision rules given by ψ_k $(k > 0, c \leq 0)$ which decides $\mu = \mu_1$, iff

$$k(1 + 1/a_1)^{-1}\|X - X_1\|_S^2 < (1 + 1/a_2)^{-1}\|X - X_2\|_S^2 + c, \tag{1.6}$$

where $\|X\|_S^2 = X'S^{-1}X$. For a rule φ, define

$$P_i(\varphi) = \Pr[\varphi \text{ decides } \mu = \mu_i | \mu = \mu_i], \ i = 1, 2. \tag{1.7}$$

In Section 3, we obtain the following results from Theorems 1 and 2.

Theorem 3.

When $\Sigma = I_p$,

(a) $P_1(\varphi_k)$ is a strictly increasing function of $\|\mu_1 - \mu_2\|$, if

$$(1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1} \leq k, \tag{1.8}$$

and

(b) both $P_1(\varphi_k)$ and $P_2(\varphi_k)$ are strictly increasing functions of $\|\mu_1 - \mu_2\|$, if

$$c = 0, (1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1} \leq k \leq (1 + 1/a_1)(1 + 1/a_2). \tag{1.9}$$

Theorem 4.

(a) If (1.8) holds, $P_1(\psi_k)$ is a strictly increasing function of $\|\mu_1 - \mu_2\|_\Sigma$.

(b) If (1.9) holds, both $P_1(\psi_k)$ and $P_2(\psi_k)$ are strictly increasing functions of $\|\mu_1 - \mu_2\|_\Sigma$.

It may be noted that for $k = 1$, and $k = (1 + 1/a_1)(1 + 1/a_2)^{-1}$, $c=0$, the rule φ_k becomes a likelihood-ratio rule [1, pp.141-142] and the minimum distance rule, respectively; the same is true for ψ_k. For these values of k, $P_1(\varphi_k)$ and $P_1(\psi_k)$ were studied by John [4, 5] and Sitgreaves [8]; but the expressions obtained by them are too complicated to obtain results like Theorems 3 and 4.

In Section 4, we have studied some estimates of the probability of correct classification of the minimum distance rule for classification into $N_p(\mu_1, \Sigma)$ or $N_p(\mu_2, \Sigma)$. Consider three sets of random samples (Z), $(X_1, \ldots, X_n)$, and $(Y_1, \ldots, Y_m)$ from $N_p(\mu, \Sigma)$, $N_p(\mu_1, \Sigma)$, and $N_p(\mu_2, \Sigma)$, respectively.

When μ_1, μ_2 and Σ are known, the minimum distance rule, given by δ, decides $\mu = \mu_1$, iff

$$(1.10) \qquad \|Z - \mu_1\|_\Sigma < \|Z - \mu_2\|_\Sigma .$$

The probabilities of correct classification (PCC) of the rule δ are given by

$$(1.11) \qquad P_1(\delta) = P_2(\delta) = \Phi(\Delta/2),$$

where

$$\Phi(u) = \int_{-\infty}^{u} (e^{-t^2/2}/\sqrt{2\pi})dt;$$

$$(1.12) \qquad \Delta = \|\mu_1 - \mu_2\|_\Sigma .$$

We shall now consider the rule $\hat{\delta}$, which is a "plug-in version" of the rule δ, defined as follows:

Case 1. Σ is known and taken to be I_p.

The rule $\hat{\delta}$ decides $\mu = \mu_1$, iff

$$(1.13) \qquad \|Z - \bar{X}\| < \|Z - \bar{Y}\|,$$

where $\bar{X}$ and $\bar{Y}$ are the means of the X-observations and the Y-observations, respectively.

Fisher [2] and Smith [7], respectively, suggested the following estimates of $P_1(\delta)$:

(1.14) $\hat{P}_1(\hat{\delta}) \equiv \Phi(\hat{\Delta}/2)$, $\hat{\Delta} = \|\bar{X} - \bar{Y}\|$,

(1.15) $c_1(\hat{\delta})$ = the proportion of X-observations correctly classified by $\hat{\delta}$.

It follows from Sorum [9] that

(1.16) $E[\hat{P}_1(\hat{\delta})] < E[c_1(\hat{\delta})] = E[\Phi\{\|X-Y\|/2(1 - 1/n)^{\frac{1}{2}}\}]$,

and from Hills [3]

(1.17) $E[P_1(\hat{\delta})] < P_1(\delta)$,

when $n = m$. We shall show that

(1.18) $P_1(\delta) \leq \Phi[\|\mu_1 - \mu_2\|/2(1 - \frac{1}{2}n)^{\frac{1}{2}}] \leq E[c_1(\hat{\delta})]$

when $m = n$, which generalizes the corresponding result of Hills [3] for $p = 1$. The inequalities in (1.18) is strict if $\mu_1 \neq \mu_2$.

<u>Case 2</u>. Σ is unknown.

Here we redefine the rule $\hat{\delta}$ as follows: $\hat{\delta}$ decides $\mu = \mu_1$, iff

(1.19) $\|Z - \bar{X}\|_S < \|Z - \bar{Y}\|_S$.

Note that for $p = 1$, (1.19) is the same as (1.13). Sorum [10], Lachenbruch and Mickey [6] obtained some numerical results regarding the performance of this rule $\hat{\delta}$. It follows from Hills [3] that

(1.20) $P_1(\hat{\delta}) < P_1(\delta)$,

if $n = m$. We shall show that

(1.21) $P_1(\delta) < E[c_1(\hat{\delta})]$

when $n = m$, where $c_1(\hat{\delta})$ is defined as in (1.15) with the rule $\hat{\delta}$ in (1.19).

2. Probability Inequalities.

Proof of Theorem 1.

By taking an orthogonal matrix L with its first row proportional to μ' and transforming $X \to LX$, $Y \to LY$, it can be seen that G involves μ only through $\mu'\mu$.

We shall first express the region $X'Y < c$ as a region R in the space of some properly defined random variables W and $V'V$, where the distribution of W does not involve μ, $k(V'V) \sim \chi_p^2(r(\mu'\mu))$, and W and $V'V$ are independent. Let $R(\lambda)$ be the section of R in the W-space for fixed $\sqrt{V'V} = \lambda$. Next we shall show that for certain values of a, ρ and c, the region $R(\lambda)$ increases as λ increases, and in certain cases $R(\lambda)$ strictly increases with λ with positive probability. These facts are enough to prove the desired result, since the density of $V'V$ has the strict monotone likelihood-ratio property in $\mu'\mu$.

Case (i).

Suppose $-1 < \rho < 1$, $-1 \leq a \leq 0$. Define

(2.1) $$U = aX - Y, \; V = X - bY,$$

where

(2.2) $$b = (\rho - a)/(1-\rho a) .$$

Then U and V are independent, and

(2.3) $$U \sim N_p(0, (a^2+1-2a\rho)I_p), \; V \sim N_p((1-ab)\mu, (b^2+1-2b\rho)I_p) .$$

Note that $ab - 1 \neq 0$. Now

(2.4) $$X'Y < c \Leftrightarrow bU'U + aV'V - (ab+1)U'V < c(ab-1)^2 .$$

Let L be a $p \times p$ orthogonal matrix with its first row proportional to V'. Define

(2.5) $$LU = W = (W_1, \ldots, W_p)' .$$

Then V and W are independent, and the distribution of W is the same as that of U. The region in (2.4) is equivalent to

$$bW'W + aV'V - (ab+1)W_1 \sqrt{V'V} < c(ab-1)^2 . \tag{2.6}$$

Let $R(\lambda)$ be the section of the region in (2.6) in W-space for fixed $\sqrt{V'V} = \lambda$, and let the conditional probability of the region, given $\sqrt{V'V} = \lambda$, be $g(\lambda)$.

If $ab + 1 = 0$, i.e., $a = -1$, (2.6) reduces to

$$W'W < V'V + 4c . \tag{2.7}$$

Now $g(\lambda) = 0$, if $\lambda^2 + 4c \leq 0$, and $g(\lambda)$ strictly increases in λ for $\lambda^2 + 4c > 0$.

If $b > 0 \Leftrightarrow a < \rho$, (2.6) can be expressed as

$$[bW'W - c(ab-1)^2](V'V)^{-\frac{1}{2}} + a(V'V)^{\frac{1}{2}} < (ab+1)W_1 \tag{2.8}$$

$$\Leftrightarrow b[W_1 - (ab+1)\sqrt{V'V}/2b]^2 + \sum_{i=2}^{p} W_i^2 < (ab-1)^2[c+V'V/4b]. \tag{2.9}$$

Now $g(\lambda) = 0$ unless $\lambda^2 + 4bc > 0$. It follows from (2.8) that when $g(\lambda) > 0$, $a \neq -1$, $R(\lambda)$, as well as, $g(\lambda)$ strictly increase in λ if $c \leq 0$.

If $b = 0 \Leftrightarrow a = \rho$, (2.8) reduces to

$$-c(V'V)^{-\frac{1}{2}} + a(V'V)^{\frac{1}{2}} < W_1 . \tag{2.10}$$

Now $g(\lambda) > 0$ for all $\lambda > 0$, and $g(\lambda)$ strictly increases in λ when $c \leq 0$ unless $a = c = 0$, in which case $P[X'Y < c] = 1/2$. The result now follows by noting that the density of

$$V'V/(b^2+1-2b\rho) \sim \chi_p^2[(ab-1)^2\mu'\mu/(b^2+1-2b\rho)] \tag{2.11}$$

which is the non-central chi-square variate and its density has the strict monotone likelihood-ratio property in $\mu'\mu$.

Case (ii).

This follows from (i) after noting that for $a \neq 0$,

(2.12) $$G(\mu;\ a,\rho,c) = G(a\mu;\ 1/a,\rho,c)\ .$$

Case (iii).

Suppose $\rho = +1$, $a \leq 0$. In this case we can express

(2.13) $$X - \mu = Y - a\mu = U,$$

where $U \sim N_p(0, I_p)$. Then

(2.14) $$X'Y < c \Leftrightarrow U'U + a\mu'\mu + (1+a)U'\mu < c$$

(2.15) $$\Leftrightarrow Z'Z + a\mu'\mu - c < -(1+a)Z_1\sqrt{\mu'\mu}$$

(2.16) $$\Leftrightarrow [Z_1 + (1+a)\sqrt{\mu'\mu}/2]^2 + \sum_{i=2}^{p} Z_i^{\,2} < c + (1-a)^2\mu'\mu/4$$

where $(Z_1,\ldots, Z_p)' = Z = LU$, and L is a $p \times p$ orthogonal matrix with its first row proportional to μ'. If $\mu = 0$, $c \leq 0$, $G = 0$. If $a = -1$, G increases strictly with $\mu'\mu$, unless $c + \mu'\mu < 0$, in which case $G = 0$. If $a \neq -1$, then it follows from (2.15) and (2.16) that G increases strictly with $\mu'\mu$ when $c \leq 0$, unless $c + (1-a)^2\mu'\mu/4 < 0$, in which case $G = 0$.

Suppose now $\rho = -1$. Then we can write

(2.17) $$X - \mu = -(Y-a\mu) = U,$$

where $U \sim N_p(0, I_p)$. Then, for $a = -1$

(2.18) $$X'Y < c \Leftrightarrow \|U + \mu\|^2 > -c\ .$$

Note that $\|U + \mu\|^2 \sim \chi_p^{\,2}(\mu'\mu)$. Thus G strictly increases with $\mu'\mu$, if $c < 0$ and for $c \geq 0$, $G = 1$.

Proof of Theorem 2.

Here the basic arguments are exactly the same as in the proof of Theorem 1. We shall express $X'S^{-1}Y < c$ in a suitable form so that the arguments are applicable.

Case (i). $-1 < \rho < 1$, $-1 \leq a \leq 0$.

Define U and V as in (2.1). Consider a $p \times p$ orthogonal matrix L with its first row proportional to V'. Define

$$U^* = LU, \quad S^* = LSL'. \tag{2.19}$$

Then the distribution of (U^*, S^*, V) is the same as that of (U, S, V). Also note that

$$U'S^{-1}U = U^{*'}S^{*-1}U^*, \quad V'S^{-1}V = S^{*11}V'V, \quad U'S^{-1}V = (U^{*'}S^{*-1}e)\sqrt{V'V}, \tag{2.20}$$

where $S^{*-1} = [S^{*ij}]$, $e' = (1,0,\ldots, 0) : p \times 1$. Let $S^{*-1} = TT'$, where T is lower triangular matrix $[t_{ij}]$, and define $W = T'U^* = (W_1,\ldots, W_p)'$. Then

$$U^{*'}S^{*-1}U^* = W'W, \quad S^{*11} = t_{11}^2, \quad U^{*'}S^{*-1}e = t_{11}W_1. \tag{2.21}$$

Now, it can be seen from (2.4), (2.20), and (2.21) that

$$X'S^{-1}Y < c \Leftrightarrow bW'W + at_{11}^2 V'V - (ab+1)t_{11}W_1\sqrt{V'V} < c(ab-1)^2. \tag{2.22}$$

The above form is similar to (2.6). Let the conditional probability of the region in (2.22), given $\sqrt{V'V} = \lambda$, $T = t$, be $h(\lambda, t)$. To get the desired result we argue exactly as in the proof of Theorem 1 with $h(\lambda, t)$, for fixed t, instead of $g(\lambda)$.

Case (ii).

This follows from (i) after noting that for $a \neq 0$

$$H(\mu; a,\rho,c) = H(a\mu; 1/a,\rho,c). \tag{2.23}$$

Case (iii).

Suppose $\rho = +1$, $a \leq 0$. Define U as in (2.13). Let L be a $p \times p$ orthogonal matrix with its first row proportional to μ'. Define

$$U^* = LU, \quad S^* = LSL'. \tag{2.24}$$

Then (U^*, S^*) is distributed as (U, S), and

(2.24) $\quad U'S^{-1}U = U^{*'}S^{*-1}U^{*},\ U'S^{-1}\mu = (U^{*'}S^{*-1}e)(\mu'\mu)^{\frac{1}{2}},\ \mu'S^{-1}\mu = s^{*11}(\mu'\mu),$

where $S^{*-1} = [s^{*ij}]$, $e = (1,0,\ldots,0) : p \times 1$. Let $S^{*-1} = TT'$, where $T = [t_{ij}]$ is lower triangular. Define $Z = T'U^{*} = (Z_1,\ldots,Z_p)'$. Then

(2.25) $\quad U^{*'}S^{*-1}U^{*} = Z'Z,\quad s^{*11} = t_{11}^2\ ,\quad U^{*'}S^{*-1}e = t_{11}Z_1\ .$

From (2.14), (2.24), and (2.25), we get

(2.26) $\quad X'S^{-1}Y < c \Leftrightarrow Z'Z + at_{11}^2\ \mu'\mu + (1+a)t_{11}\sqrt{\mu'\mu} < c$

(2.27) $\quad \Leftrightarrow [Z_1 + (1+a)t_{11}\sqrt{\mu'\mu}\ /2]^2 + \sum_{i=2}^{p} Z_i{}^2 < c + (1-a)^2 t_{11}^2\mu'\mu/4\ .$

Let $h(\Delta, t)$ be the conditional probability of the above region, given $T = t$, and $\sqrt{\mu'\mu} = \Delta$. Clearly $h(\Delta, t) = 0$, iff

$$c + (1-a)^2 t_{11}\ \Delta^2/4 \leq 0,$$

and otherwise $h(\Delta, t)$ strictly increases with Δ if $a = -1$ or $a \neq -1$ and $c \leq 0$.

Suppose $\rho = -1$, $a = -1$. Define U as in (2.17). Now

$$X'S^{-1}Y < c \Leftrightarrow \|U + \mu\|_S^2 > -c\ .$$

Note that the density of $\|U + \mu\|_S^2$ has the monotone likelihood-ratio property in $\mu'\mu$. Thus H strictly increases with $\mu'\mu$, if $c < 0$ and for $c \geq 0$, $H = 1$.

3. Monotonicity of PCC.

Proof of Theorem 3.

Define

$$(3.1)\qquad U = [\sqrt{\lambda}(X - X_1) - (X - X_2)]/\tau_1$$

$$(3.2)\qquad V = [\sqrt{\lambda}(X - X_1) + (X - X_2)]/\tau_2$$

where

$$(3.3)\qquad \lambda = k(1 + 1/a_1)^{-1}(1 + 1/a_2)$$

$$(3.4)\qquad \tau_1^{2} = \lambda(1 + 1/a_1) + (1 + 1/a_2) - 2\sqrt{\lambda}\ ,\quad \tau_1 > 0$$

$$(3.5)\qquad \tau_2^{2} = \lambda(1 + 1/a_1) + (1 + 1/a_2) + 2\sqrt{\lambda}\ ,\quad \tau_2 > 0.$$

Then (1.5) can be expressed as

$$(3.6)\qquad U'V < c^*, \text{ where } c^* = c(1+ 1/a_2).$$

When $\mu = \mu_1$, U and V are jointly normally distributed as 2p-variate normal distribution with mean vectors $-\mu/\tau_1$ and μ/τ_2, respectively, where $\mu = \mu_1 - \mu_2$, and the covariance matrix

$$(3.7)\qquad \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \otimes I_p\ ,$$

where

$$(3.8)\qquad \rho = [\lambda(1 + 1/a_1) - (1 + 1/a_2)]/\tau_1\tau_2\ .$$

Let $a = -\tau_2/\tau_1$. Then $a < -1$. Note now

$$(3.9)\qquad a\rho \leq 1 \Leftrightarrow \lambda \geq (1 + 1/a_1)^{-2}$$
$$\Leftrightarrow k \geq (1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1}\ .$$

Theorem 3(a) follows now from Theorem 1(ii). The part (b) follows from (a) after replacing k by $1/k$.

Proof of Theorem 4.

Use Theorem 2 and the proof of Theorem 3.

4. Estimation of PCC.

First we shall prove (1.18). Let a be a vector such that $a'a = 1$. From Hills [3], we get

$$(4.1) \qquad \Phi[|a'(\mu_1 - \mu_2)|/2] \le \Phi[|a'(\mu_1 - \mu_2)|/2(1-\tfrac{1}{2}n)^{\frac{1}{2}}]$$
$$\le E\{\Phi[|a'(\bar{X} - \bar{Y})|/2(1-1/n)^{\frac{1}{2}}]\} .$$

However,

$$(4.2) \qquad E\{\Phi[|a'(\bar{X} - \bar{Y})|/2(1-1/n)^{\frac{1}{2}}]\} \le E\{\Phi[\sup_{a'a=1} |a'(\bar{X} - \bar{Y})|/2(1-1/n)^{\frac{1}{2}}]\}$$
$$= E\{\Phi[\|\bar{X} - \bar{Y}\|/2(1-1/n)^{\frac{1}{2}}]\} = E[c_1(\hat{\delta})] .$$

Also

$$(4.3) \qquad \sup_{a'a=1} \Phi[|a'(\mu_1 - \mu_2)|/2(1-1/2n)^{\frac{1}{2}}] = \Phi[\|\mu_1 - \mu_2\|/2(1-1/2n)^{\frac{1}{2}}].$$

From these relations we get (1.18).

From the fact that $\Phi(t)$ is a concave function when $t \ge 0$, we get

$$(4.4) \qquad E[c_1(\hat{\delta})] = E[\Phi\{\|\bar{X} - \bar{Y}\|/2(1-1/n)^{\frac{1}{2}}\}] < \Phi\{E\|\bar{X} - \bar{Y}\|/2(1-1/n)^{\frac{1}{2}}\}$$
$$< \Phi\{\sqrt{E\|\bar{X} - \bar{Y}\|^2} \;/2(1-1/n)^{\frac{1}{2}}\}$$
$$= \Phi\{\sqrt{\|\mu_1 - \mu_2\|^2 + p(1/n + 1/m)} \;/\; 2(1-1/n)^{\frac{1}{2}}\}.$$

The question of getting upper bounds for $P_1(\hat{\delta})$ (when $n \neq m$) and $E\hat{P}_1(\hat{\delta})$ may be partially resolved as follows. Consider the validity of the following relation. For $U_p \sim N_p(EU_p, I_p)$,

$$(4.5) \qquad E[\Phi(a\|U_p\|)] \le \Phi[b\|EU_p\|],$$

<u>for all p</u>, where $a > 0$, $b > 0$ are constants independent of p. Define $W'_{p+q} = (U'_p\ V'_q)$, where U_p, V_q are independent and $V_q \sim N_q(0, I_q)$. Clearly,

$$E[\Phi(a\|U_p\|)] < E[\Phi(a\|W_{p+q}\|)]$$

for $q > 0$. Note that $\|EU_p\| = \|EW_{p+q}\|$ and

$$E[\Phi(a\|W_{p+q}\|)] \to 1 \quad \text{as} \quad q \to \infty.$$

Thus (4.5) is not true. In the left hand side of (4.5) Φ may be replaced by any strictly increasing function and in the right hand side Φ may be replaced by any function; still (4.5) is not true. From the table in Hills [3], it is found that for $p = 1$

$$E[\hat{P}_1(\hat{\delta})] < P_1(\delta),$$

but this cannot be true for all p. Let us consider now $P_1(\hat{\delta})$ when $n \neq m$, and study the validity of the relation

$$P_1(\hat{\delta}) \leq g(\|E(\overline{X} - \overline{Y})\|) \tag{4.6}$$

for all p, for some function g. It can be seen as above, that for fixed $\|E(\overline{X} - \overline{Y})\|$,

$$P_1(\hat{\delta}) \to 1$$

as $p \to \infty$, if $n > m$. Thus (4.6) is not true.

Let us now prove (1.21). Without loss of generality we shall assume that $\Sigma = I_p$.

$$\begin{aligned} E[c_1(\hat{\delta})] &= P[(\overline{Y} - \overline{X})'S^{-1}(X_1 - (\overline{X} + \overline{Y})/2) < 0] \\ &= P[W < \|\overline{X} - \overline{Y}\|_S \,/2(1-1/n)^{\frac{1}{2}}], \end{aligned} \tag{4.7}$$

where

$$W = (\overline{Y} - \overline{X})'S^{-1}(X_1 - \overline{X})(1-1/n)^{-\frac{1}{2}}/\|\overline{X} - \overline{Y}\|_S \ . \tag{4.8}$$

It can be shown that W is independent of $\|\overline{X} - \overline{Y}\|_S$, and the distribution W is the same as that of

$$W_1/(W_1^2 + W_2^2)^{\frac{1}{2}}$$

where W_1 and W_2 are independent, $W_1 \sim N(0, 1)$, $W_2^2 \sim \chi^2_{f-1}$, and $f = m + n - 2$.

We shall use the fact that the c.d.f. of W, given by F, is independent of p. Thus

$$(4.9) \qquad E[c_1(\hat{\delta})] = E[F\{\|\bar{X} - \bar{Y}\|_S /2(1-1/n)^{\frac{1}{2}}\}] .$$

Let a be a $p \times 1$ vector such that $a'a = 1$. From Hills [3] we get

$$(4.10) \qquad E[F\{(|a'(\bar{X} - \bar{Y})|/(a'Sa)^{\frac{1}{2}})/2(1-1/n)^{\frac{1}{2}}\}] = E[\Phi\{|a'(\bar{X} - \bar{Y})|/2(1-1/n)^{\frac{1}{2}}\}]$$

$$> \Phi[|a'(\mu_1 - \mu_2)|/2(1-1/2n)^{\frac{1}{2}}].$$

The relation (4.10) easily yields

$$(4.11) \qquad E[c_1(\hat{\delta})] \geq \Phi[\|\mu_1 - \mu_2\|/2(1-1/2n)^{\frac{1}{2}}]$$

$$\geq \Phi[\|\mu_1 - \mu_2\|/2] = P_1(\delta),$$

the inequality being strict if $\mu_1 \neq \mu_2$.

References.

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. J. Wiley & Sons, New York.

[2] Fisher, R. A. (1936). Use of multiple measurements in taxonomic problems. Ann. Eugen. Lond. 7 179-188.

[3] Hills, M. (1966). Allocation rules and their error rates. Jour. Royal Statist. Soc. Ser. B 28 1-31.

[4] John, S. (1960). On some classification statistics. Sankhyā 22 309-316.

[5] John, S. (1961). Errors in discrimination. Ann. Math. Statist. 32 1125-1144.

[6] Lachenbruch, P. A. and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. Technometrics 10 1-11.

[7] Smith, C. A. B. (1947). Some examples of discrimination. Ann. Eugen. Lond. 13 272-282.

[8] Sitgreaves, R. (1961). Some results on the distribution of the W-classification statistic. Studies in Item Analysis and Prediction, ed. H. Solomon. Stanford University Press, Stanford.

[9] Sorum, M. J. (1971). Estimating the conditional probability of misclassification. Technometrics 13 333-343.

[10] Sorum, M. J. (1972). Three probabilities of misclassification. Technometrics 14 309-316.